ELLSBERG

# IV

## POWER ECONOMICS

The economic man is a natural target for the blackmailer; for blackmail requires a rational victim. Not "behavior," in general, but rational choice is swayed by blackmail: choice among alternative actions, insofar as it is determined by expectations and preferences. For blackmail operates upon expectations: specifically, the victim's expectations of the blackmailer's own actions.

"consistent" not [illegible] [illegible], [illegible], "[illegible]" ([illegible]) (a "[illegible]" open to the [illegible]

For these expectations to influence him, the outcomes of the victim's acts must depend upon the choices by the blackmailer. Thus, the possible actions open to the two define, in part, a game (since, presumably, it is likewise true that the victim's choices "make a difference" to the blackmailer). Still further conditions--beyond rationality and dependence of outcome--hedge the victim's predicament, the degree of his vulnerability and the nature of the blackmailer's problem. Rather than labor these conditions abstractly, let us examine them in the context of a simple game.*

*(a 'cooperative non-zero-sum two-person minorant game')

Let us suppose that the victim has only two alternative actions, or strategies, and that he must choose first (this is a "minorant" game for the victim). The blackmailer, too, has only two possible alternatives; he chooses second, in full knowledge of the victim's decision. For convenience, let us say that "I" am the blackmailer, "you" the victim. In the matrix below, your decision is represented as the choice of a column, mine as the (subsequent) choice of a row. The number at the intersection of a row and column does not, in this example, represent the standard game payoff; it indicates only your ordinal preference for the outcome determined by this joint choice of strategies, "0" corresponding to the outcome you least prefer. As ordinal utilities, these numbers are determined only up to a

monotonic transformation.

(1)

| | | Victim I | Victim II |
|---|---|---|---|
| Blackmailer | A | 1 | 2 |
| | B | 1 | 0 |

My payoffs are not shown; but let us suppose that I would prefer you to choose your strategy I. To begin with, however, (i.e., in the absence of my "blackmail") I expect you to choose strategy II: because, let us say, you are certain (for some reason) that I will respond with row A, giving you your best outcome. My problem as a blackmailer is to induce you to choose <u>I</u> by magnifying your expectation that I would respond to strategy <u>II</u> by choosing B (resulti in your worst outcome).

Thus, I must be able somehow to communicate with you, directly or indirectly; more specifically, I must be able to take some actions (quite outside the possibilities represented by strategies A and B) designed to <u>change</u> your expectations of my responses. In other words, this must be a "cooperative" game, with communication allowed. However, in this chapter we will simplify the argument by assuming the communication is one-way, possible only from me to you; you have no means for operating upon <u>my</u> expectations.

We have already mentioned that your outcomes must be dependent on my choices. This rules out matrices like the following:

(2)

| | I | II |
|---|---|---|
| A | 1 | 1 |
| B | 1 | 1 |

(3)

| | I | II |
|---|---|---|
| A | 0 | 2 |
| B | 0 | 2 |

Clearly, in these cases I could not influence your behavior by threats. But this might also be true even though I could affect your outcome;

I still might not ~~infx~~ be able to influence your choice. Consider the following case:

(4)

| | I | II |
|---|---|---|
| A | 0 | 3 |
| B | 1 | 2 |

Here you would, indeed, prefer that I should choose A rather than B; but your choice will be quite unaffected by your expectation of my choice, whatever you expect me to do. You are not indifferent to my choice; yet you are invulnerable to my threats.

What puts you beyond the reach of blackmail in this case is that you have a strategy whose worst outcome is preferable to the best outcome offered by any alternative strategy. Another condition for a prospective victim, then, is that he have no such strategy. This by no means indicates that a merely dominant strategy confers invulnerability. In the following case, strategy II "dominates" strategy I; "given" that I would respond with A, you would prefer II, and likewise, "given" that I would choose B.

"I" and "II" confusing

(5)

| | I | II |
|---|---|---|
| A | 2 | 3 |
| B | 0 | 1 |

But since this is a minorant game, in which I choose second, my choices cannot be regarded as "given," independently of your choice; ~~the conditional probabilities of my choice~~ on the contrary, my choice will presumably depend on yours. Given means of communication, I can set out to convince you, say, that strategy II will be followed with B; I with A. The "dominance" of II is then quite irrelevant.

(Cf. The probab. of my choice ("external event" — in a game against Nature) in your mind is affected by what you choose (focal:)

Given, then: a) you are rational; b) your outcomes are dependent on my choices; c) the outcomes of your strategies "overlap" in preference, the minimum under each being worse than the maximum under the other; d) I have means of communication with you, and may be able to change your expectations of my behavior: you may be a candidate for blackmail. Let us return to case (1).

| | | | Comply I | Resist II |
|---|---|---|---|---|
| (1) | Accept | A | 1 | 2 |
| | Punish | B | 1 | 0 |

If I can manage to induce some uncertainty in your mind as to my response, then essentially you must choose between a <u>certain</u> outcome (with utility number 1), and a <u>gamble</u>, in which your outcome may be better or it may be worse.

The strategies have been <u>relabelled</u>, but the labels are still rather abstract. The model might correspond to a great many concrete situations. In a bargaining context, the outcomes might be terms of trade, rates of exchange, with "0" corresponding to "no deal"; Resist might correspond to <u>your</u> insistence on your own "last offer," which, if I should Accept, would give you your best outcome. If, however, I responded with Punish (the term is not perfectly suited to this particular context), the resultant "no deal" would be worse than the settlement you could have ensured by choosing Comply.

or 0 could be much worse than status quo — for both of us!

In labor negotiations, the threatened punishment for failing to Comply might be a strike or a lockout. If I were a racketeer,

I might Punish with demolition or physical violence. As an employee my threatened second strategy may be to quit; as an employer, to fire. As a rival business firm, I may threaten to undercut prices, to launch an advertising war, to withhold supplies or financing, to boycott, to deny access, to withhold some agreement, to bring suit. (Leaving the economic sphere, we could obviously apply the same simplified model to, say, negotiations between states. The issue might be control over a piece of territory, or the terms of a treaty; the strategy Punish might cover sanctions from breaking off negotiations--resulting in "no deal"--through economic pressure and propaganda, to various levels of war).

There is no necessity that the outcomes of these choices be physically quantifiable (though they might be, in the case of rates of exchange, wage rates, or profits resulting from agreement or lack of agreement). So far we have required only that you be able to compare these outcomes in terms of your own preference.

or prop- destroyed, a furnace

My problem as a blackmailer is to convince you that I am "too" likely" to respond with Punish for you to accept the risk that Resist would entail. If I could persuade you that I am certain to choose Punish, my problem ~~is~~ would be solved; but this might be impossible for me. In any case, let us say, I cannot rely on doing it. Fortunately for me, your certainty is not required. To influence you, my threatened punishment need not, in general, be certain, but only "sufficiently likely."[1]

?

1. At least, this will be true for all victims who obey the von Neumann-Morgenstern utility axiom 3:B:c, the "continuity" axiom, which states that if the individual prefers outcome A to B, B to C, and A to C, then for some probability p, $0 < p < 1$, he will ~~be indiffer between~~ still prefer the "lottery ticket" (A,C;p) ~~and~~ to the certainty of B. ~~That is, he will be indifferent between the certainty of B and~~ (a gamble which offers A with probability p and C with probability 1-p.) See von Neumann and Morgenstern, Theory of Games, New Jersey, 1947, p. 26.

lexico?

How likely must it be? There, I suggest, is a crucial question in any fruitful analysis of blackmail, and ultimately of bargaining. Yet in our discussion so far--and in any discussion of bargaining or bilateral monopoly that relies upon indifference-map, or ordinal preference data--there is no basis for answering it. Knowing preferences only, we have no grounds for inferring the risks the victim will take on the basis of these outcomes. To know the ~~it~~ expectation of punishment ~~that~~ in the victim's mind that the blackmailer must "achieve"--and hence, the ease or difficulty of achieving it, the likelihood that he will succeed, the relative effectiveness of different threat-tactics ~~and~~ or of given tactics against different victims--we must know the victim's preferences among "lottery tickets" as well as among outcomes. We must know the actual choices he would make among lottery tickets that combine these various outcomes. We must, in short, have the sort of data that may be represented by a von Neumann-Morgenstern utility function.

OR?

We could, as a matter of fact, make predictions about the victim's behavior and the blackmailer's required tactics even though the victim's risk-preferences were not orderly enough to be represent by such a utility index. But it will simplify our discussion in this chapter enormously if we assume that you, the victim, do obey the von Neumann-Morgenstern utility axioms in choosing among gambles; and that I know your von Neumann-Morgenstern utilities for the outcomes in question. This happens to be a standard assumption in game-theoretical discussions of bargaining, and it is implicit in many older discussions, such as that of Zeuthen (as we shall see); we shall relax it, in later chapters.

?

Under this assumption, we can answer very neatly the questions "How likely ~~th~~ must the threatened punishment be?" or "How sure do you have to be that I will carry out my threat before you will choose Comply?" Let us suppose (and this is a major assumption, to be discussed in detail later) that your expectations, your uncertainties or degrees of belief, concerning my choice can be represented by subjective probabilities. If you obey the utility axioms (with respect to these probabilities), there will be a unique probability p for which you will be indifferent between Comply and Resist, when you assign likelihood p to my Punish and 1-p to Accept. To find this probability--experimentally, by observation or interrogation, by inference--is to answer the above questions in a particular case. For such a probability would be a threshold. It would correspond to the maximum likelihood of Punish you would accept in choosing to Resist. If you assigned less likelihood than p to Punish, you would ~~Comply~~ Resist; if you thought me more likely than probability p to carry out my threat, you would Comply. This, then, is the credibility, the degree of belief, that I must achieve for my threat, if it is to influence you. We shall refer to this threshhold probability--which will depend in each case on the particular outcomes threatened, demanded and offered--as your critical risk. A more suggestive label might be your "willingness to Resist,"[2] if we keep in mind that this "willingness" is defined (measured) in terms of a subjective probability.

2. This corresponds directly to Zeuthen's notion, "~~willingness~~ readiness to fight" (Frederick Zeuthen, Problems of Monopoly and Economic Warfare, London, 1930, Chapter IV), which will be discussed at length below.

If we did not know your von Neumann-Morgenstern utilities, but knew that you obeyed the utility axioms, we could nevertheless set out to determine your "critical risk" experimentally, by observing your choices among various gambles (or, less reliably ~~more sloppily~~?, by asking you questions about hypothetical choices). Let us refer to the best and worst outcomes under Resist, respectively, as A and C, and the consequence of Comply as B (A,B,C in order of decreasing preference). Our problem would "simply" be to discover a p such that you were indifferent between B and (C,A;p) ~~(A,C;p)~~, where the latter expression denotes a lottery ticket offering outcome ~~A~~ C with probability p and A with probability 1-p.

"objective" (or if one knew your subjective probs...)

Incidentally, having done so we could proceed to assign these outcomes von Neumann-Morgenstern utilities. Suppose the threshhold probability of Punish that would make you indifferent between your two strategies were found to be 1/10 ~~9/10~~.[3] Then, assigning arbitrarily

3. Any such precision would be empirically spurious; the substance of the argument below will not at all depend upon attaining any such precise results, either in the payoffs or the expectations. However, it simplifies the discussion considerably to imagine such hypothetical results; and the results suggested can be compared immediately with those in other discussions, particularly game-theoretical, which typically assume such data as a matter of course. I shall, however, proceed below not only to "coarsen" such data in general but to question in some cases its conceptual basis.

the utility numbers 100 to A and 0 to B[4], the utility number to

4. In assigning two numbers arbitrarily, we fix the origin and scale of our utility index.

be assigned to B is determined by the equation:

$$u(B) = 9/10 \cdot 100 + 1/10 \cdot 0 = 90$$

Your payoffs, now in the form of von Neumann-Morgenstern utilities, would then appear:

(6) $$\begin{array}{ccc} & \mathrm{I} & \mathrm{II} \\ A & u(B) & u(A) \\ B & u(B) & u(C) \end{array} \quad \backsim \quad \begin{array}{ccc} & \mathrm{I} & \mathrm{II} \\ A & 90 & 100 \\ B & 90 & 0 \end{array}$$

To present these numbers as your von Neumann-Morgenstern utilities is to say that we <u>know</u> (on some basis) that you are, in fact, indifferent between Comply and Resist if you believe that I am 1/10 likely to choose Punish. Through observation, or interrogation, or inference from other choices (or, to be honest, through guess or hypothesis or empathy) we know that you would prefer the consequence of Comply to a gamble that offered you the two possible outcomes of Resist each with .5 probability. If this were not so, these particular utility numbers would simply be incorrect. Similarly, if these numbers are appropriate, they indicate that (we know) you would prefer a gamble offering the best outcome under Resist with

or even .91+ .99 probability (and the worst with .01 probability) to the certain outcome of Comply.

From now on we shall assume the von Neumann-Morgenstern utilities of the players <u>given</u>; which is to say, we assume that we possess this sort of knowledge about the players' preferences among lotteries. And when the von Neumann-Morgenstern utility payoffs are given, in a 2x2 matrix such as this one, the critical risk can simply be be computed from them; it is implicit in the payoffs. In a matrix of this sort, the threshhold probability <u>p</u> which makes the player indifferent between his two strategies is determined by the

equation:

$$u(B) = (1-p)\cdot u(A) + p\cdot u(C)$$

or, in this case:

$$90 = (1-p)\cdot 100 + p\cdot 0, \qquad p = 1/10$$

The critical risk is thus a function of the victim's payoffs, when these payoffs are in the form of von Neumann-Morgenstern utilities (i.e., when the victim's preferences are known among lotteries offering the outcomes in question as prizes, with varying probabilities). Note that it depends upon the entire structure of payoffs. The question, "How sure does the victim have to be?" (that the blackmailer will carry out his threat, before the victim chooses to comply) cannot be answered simply on knowledge of a single outcome, such as the punishment outcome; the critical risk depends upon the relationships among all the possible outcomes: the punishment outcome, the outcome the victim hopes to achieve, the outcome offered by the blackmailer.

We can now imagine you making your choice by comparing your actual prob. expectation that I will carry out my threat if you Resist--what we might call your actual risk of punishment--with your critical risk: your threshhold probability determined by your payoffs your "willingness to Resist." If you decide that the actual risk is higher than the critical level, you Comply; otherwise you resist my threats, refusing to comply. One advantage of this formulation is that to predict your choice (or, for that matter, for you to reach your decision) it is not necessary to know your expectations

precisely. It is only necessary to estimate your expectation of punishment (under Resist) as greater or less than your critical risk. If your critical risk is .25, we need discover only whether you assign more or less than .25 probability to my carrying out my threat, to predict your choice.

But we can no longer postpone facing the question that aims at the center of this analysis: Is it meaningful at all to speak of your "assigning" probabilities to my actions, to represent your expectations in terms of probabilities? What is the meaning, if any, to the statement that you assign more, or less, than .5 probability to one of my choices?

First, these statements do not imply that my choice is in any sense randomly determined, or that you think it is. They do not even imply that my decision is subject to some rationally chosen indeterminacy, as exemplified by a "mixed strategy." These "probabilities" reflect nothing more than your uncertainty as to what I will do. I may be able to predict correctly, with certainty, what I will do; you, we assume, cannot. It is proposed that your own subjective "degrees of belief" in various hypotheses as to my response can be represented by numbers having the characteristics of probabilities. This assumption, as we shall interpret it, has a good deal of empirical content; indeed, it is far too restrictive for a realistic general theory. Later we shall consider important cases in which it does not seem appropriate. However, the conditions under which it is valid seem likely to be satisfied in a significant number of cases.

It was Frank Ramsey, in his essay on "Truth and Probability" written in 1926,[5] who first suggested an empirical content (of the

5. Frank Ramsey, The Foundations of Mathematics, London, 1931

sort we shall consider) for statements involving subjective probabilities. In specifying operations for the measurement of these probabilities and empirical results which would falsify or confirm estimates of them, he gave operational meaning to assertions concerning their magnitude.

In Ramsey's view, "the degree of a belief is...the extent to which we are prepared to act upon it,"[6] and he interpreted this

6. Op. cit., p. 169.

roughly as "the extent to which we are prepared to bet on it." The actions which reveal (measure) our degrees of belief in the likelihood of certain events are choices among actions whose consequences will depend upon the occurence of one or another of these events. When the events, hence the consequences that depend on them, are uncertain, such choices have the character of gambling behavior: of choosing between various bets. "The probability of 1/3 is clearly related to the kind of belief which would lead to a bet of 2 to 1."[7] This

7. Ibid., p. 171.

is an imprecise statement (it becomes plausible only in connection with a measurement of the stakes involved in terms of "utility), which I cite merely to show the general nature of Ramsey's approach.

The application of this approach, divorced from utility considerations, appears most clearly in the case of a degree of belief of ½. This emerges as one's degree of belief in a proposition or

event which one is equally willing to bet on or against. Consider the following choice, where the rows represent actions, the columns represent the occurrence of an event E or of its complement $\bar{E}$ ("not-E"), and the ~~numbers are money payoffs~~ and the chooser prefers outcome a to outcome b.

| | $E$ | $\bar{E}$ |
|---|---|---|
| I | a | b |
| II | b | a |

Since the set of possible payoffs ~~to~~ is the same ~~action~~ for each action, any preference for one action as opposed to the other ~~alternative~~ must reflect the likelihoods which the actor assigns to the possible events. In Ramsey's approach (and in the subsequent, similar analyses by Savage, Suppes and Luce), if a person is indifferent between the two actions (for any a, b, with a preferred to b), it is inferred that he regards the events E and not-E as "equally likely." He assigns them equal "degrees of belief," and if he satisfies certain other conditions so that his degrees of belief may be represented as probabilities, his subjective probability for the event E will be represented as $\frac{1}{2}$.

If, for any a, b such that he prefers a to b, the actor prefers action I to action II, we infer that he regards E as more likely than not-E;[8] i.e., E is "more likely than not" to occur. L.J. Savage

8. This is not equivalent to finding E "more probable" than not-E; the pattern of choices must satisfy certain axioms before we are justified in representing his expectations as "probabilities." The nature of these restrictions and the likelihood that they will be satisfied will be discussed later. However, in this discussion we will assume that they do apply, so that the subject's choice do reveal probabilities, at least in a qualitative sense.

has shown[9] that if a person's pattern of choices among options of this general sort obeys certain restrictions--which he specifies in the form of axioms--then it will be possible to infer from these choices both his "personal probabilities" for the events in question and his "von Neumann-Morgenstern utilities" for the payoffs. His choices among "lottery tickets" involving these payoffs and events can then be described as maximizing the mathematical expectationsx of utility, on the basis of these inferred utilities and probabilities.

One method of approach would be to discover first an event which the person regarded as "as likely as not": i.e., one which could be assigned .5 probability. Observing his choices among lottery tickets employing this event with various payoffs, we could assign utility numbers to certain of these payoffs. On the basis of these utilities we could infer, by his choices among other lottery tickets, probabilit for other events (not = .5) and utilities for other payoffs.

For example, suppose we assign to outcomes a and b the arbitrary utility numbers 100 and 0. Suppose that E proves to be an event such that our subject is indifferent between actions I and II in the following choice:

| | $E$ | $\bar{E}$ |
|---|---|---|
| I | 100 | 0 |
| II | 0 | 100 |

We then assign to E the probability ½. Now we look for an outcome x lying between a and b in the subject's preferences, such that he is indifferent between options III and IV in the following choice:

9. Leonard J. Savage, The Foundations of Statistics, New York, 1954.

| | $E$ | $\tilde{E}$ |
|---|---|---|
| I | 100 | 0 |
| II | $x$ | $x$ |

To outcome *x* we can assign the utility number 50, satisfying the equation: $u(\underline{x}) = 1/2\cdot 0 + 1/2\cdot 100$

Similarly, we might find an outcome *y* indifferent to a 50-50 chance of *b* or *x*; to this we would assign the number 25. Then, if the subject proved to be indifferent between the options I and II in the following choice (involving the new event N), we would assign to N the subjective probability 1/4, satisfying the equation:

$$25 = p_N \cdot 100 + (1-p_N)\cdot 0$$

| | $N$ | $\tilde{N}$ |
|---|---|---|
| I | 100 | 0 |
| II | 25 | 25 |

If, contrary to our hypothesis, our subject did not obey the Savage axioms in his choices, he would short-circuit or invalidate this procedure in one way or another. For example, there might exist no outcome *x* or event N such that the subject would act in the way postulated. In any case, his choices will never be so ordered and consistent that a really precise measurement of utilities or probabilities is possible. Yet I suspect that in many cases the axioms will hold at least approximately; gross differences, at least, will be distinguishable in the data, and even these crude "measurements" may be better than none.

The game-theorist who assumes that payoffs are given in the form of von Neumann-Morgenstern utilities--hence, that the player obeys the von Neumann-Morgenstern utility axioms (which specify his choices among lottery tickets, with known probabilities attached to the prizes)--may find it easy to assume further that the player satisfies the Savage axioms in "games against Nature," so that it would be possible to infer his subjective probabilities for "Nature's strategies." But most game-theorists would shy away from a similar assumption for a game against a presumably reasonable opponent.

It is true that the Savage axioms are unlikely[10] to apply to the

10. Such statements can be taken from now on to reflect ~~my own~~ the writer's subjective probabilities.

player's choices <u>within the game</u>. If the payoffs to the strategies were changed,[11] the player's preferences and choices would be likely

11. In strict usage, the payoffs go to define the game, and if they change, the "game" is new; in standard game-theoretical discussion, the strategies could no longer be said to be "the same." But we can easily imagine the strategies as defined independently of payoffs--say, in terms of physical operations--and the game as defined by its strategies. Indeed, this corresponds to everyday usage. We can then discuss the effects on behavior of varying the payoffs within a "given" game.

to change in ways that violated the axioms. The Savage axioms lead to the assignment of probability numbers to events which are independent of the payoffs associated with those events under different strategies, and (as we shall discuss in more detail later) to the extent that a player's expectations of his opponent's choice can be measured at all, those expectations will <u>not</u> generally be independent of payoffs, either his own or his opponent's. Even in a non-zero-sum game (where

one player's payoffs can change independently of his opponent's), a change in one payoff for one player might change his strategic opportunities, hence his opponent's expectations, hence his own expectations of his opponent's probable behavior.

We are interested in measuring (at least hypothetically) a player's expectations of the opposing choices for given payoffs in the game. The choices that are to reveal these expectations must be in the nature of side bets on the choice that the opponent will make. ~~Thexxxsumptin~~ In effect, the player may be offered gambles with various odds on the possible strategies that the opponent may choose. My basic suggestion is that his preferences among these gambles will typically reveal that his expectations for various opposing strategies will differ from one to another; Even in a game against a reasonable and informed opponent, I think it is a fact that a player will have varying estimates of the likelihood of different choices by his opponent, that these expectations ~~ma~~ will be revealed in his preferences among side bets on those choices, and that these data are relevant and should be represented explicitly in a theory of his conflict behavior.

The assumption that these expectations can be represented as subjective probabilities goes somewhat further. It implies that his choices among these side bets will be consistent with the Savage axioms I do not think this will always be even approximately true. But the same objection applies to the validity of those axioms even in games against Nature; and I will argue later that in fact they will be applicable about as commonly in games against an opponent as in games against Nature. This argument must rest on an analysis of

the conditions under which the Savage axioms are particularly likely to be violated. At this point I merely assert that these conditions do not seem intrinsic to the basic game situation. Therefore, I think that we are justified in considering initially situations in which a player's expectations about his opponent's choice are sufficiently clear-cut that his preferences among side bets as to that choice will reveal measurable subjective probabilities.

Whether the introduction of expectation data (let alone subjective probabilities) is a step forward or backward from the initial corpus of game-theory is a question that may concern certified game-theorists. (It is not my concern, but the point is worthy of brief comment). It was, indeed, one of the starting-points of von Neumann and Morgenstern that a player could not reasonably assign probabilities to his opponent's choices. But it was the notion of <u>objective</u> probabilities that these authors were rejecting; they criticized the practice of treating the behavior of a reasonable opponent in a conflict situation as though it were determined stochastically, in a manner independent of one's own strategic opportunities.

12. von Neumann and Morgenstern, <u>op. cit.</u>, p. 11.

As they put it, the choices by a player's opponents "cannot, from his point of view, be described by statistical assumptions...because the others are guided, just as he himself, by rational principles." It is not at all a retreat from this insight to suppose that the player's own uncertainties about their choices may be represented by subjective probabilities: a concept which had not been extensively developed at the time that von Neumann and Morgenstern wrote. It is an abstraction of the conflict situation alternative to the one

those authors chose (which amounted to the assumption that the player was in complete ignorance as to his opponent's choice, or at least, that he made no attempt to distinguish among his opponent's possibilities in terms of likelihood).

In principle, then, the player's utilities and subjective probabilities are to be determined by his choices among gambles --independent of his strategic choices in the actual context of the game--involving his opponent's possible choices and his game outcomes. Such assumptions as to data are familiar enough in discussions of bargaining, particularly the game-theoretical, but it is worthwhile to consider briefly the practical availability of such information. Suppose that in the following example outcomes are specified only in terms of money.

| | | (Comply) I | (Resist) II |
|---|---|---|---|
| (Accept) | A | \$99 | \$100 |
| (Punish) | B | \$99 | \$0 |

Suppose that ~~it~~ we have a particular player in mind. ~~In order to get an approximate estimate of his utility payoffs, will it really be necessary~~ We may know much or little about him; but the chances are that we have <u>not</u> observed his choices among gambles involving these outcomes, or, in fact, any other outcomes: and that we might have great difficulty in doing so. Is that really necessary? The crucial question is: what probability p of \$0 would make him indifferent

between the certainty of \$99 and ~~the~~ a gamble offering \$0 with probability p and \$100 with probability (1-p)? (This is, in effect, a question as to his "critical risk" in this particular case; but it serves to determine his utilities). I suspect that we would be willing to guess at this, ~~could answer this~~ roughly, without knowing a great deal about the player or havin observed any of his specifically risky behavior. I would guess that the probability in question is: ~~very~~ low. Even, very low. This is not to put it at .05 or .01; but such precision may not be necessary.

~~If intuition does not provide that answer directly, we can approach it more gradually.~~ If utility numbers of 0 and 100 are assigned arbitrarily (to fix ~~the~~ unit and origin of the utility scale) to the outcomes \$0 and \$100, then an estimate of p ~~of~~ as .05 would lead to the assignment of 95 to the outcome \$99; an estimate of p ~~of~~ as .01 would lead to the estimate: u(\$99) = 99.[13] Our

13. This would indicate that his utility function was a linear function of money outcome. If his other choices were presumed to be consistent with this, we would be assuming that he maximized the mathematical expectation of money.

conclusions about this game situation might not at all be sensitive to this amount of uncertainty. We might perfectly well be able to distinguish this situation, in terms of utility payoffs and their strategic implications, from one in which the \$99 outcome is replaced by \$1.

| | | (Comply) I | (Resist) II |
|---|---|---|---|
| (Accept) | A | \$1 | \$100 |
| (Punish) | B | \$1 | \$0 |

I do not propose that we assume in general that utility is linear on money. My point is simply that ~~we need not always rely~~

controlled, laboratory experiments are not an absolute prerequisite to the inference of utility data. Even in the case of von Neumann-Morgenstern utilities, we ~~are~~ may be entitled to infer from gross differences in "physical" outcomes, similar differences in the corresponding utilities;[14] and such inference is very common.

14. To be sure, the risk-behavior data--i.e., the choices among lottery tickets--are basic to the measurement of von Neumann-Morgenstern utilities; see "Classic and Current Notions of Measurable Utiliy," Economic Journal, September, 1954. But there will often be a reliable empirical basis for inferring these utilities, even from non-risk data.

---

In the last example: Would we predict that the player would choose Resist if he thought that the "blackmailer" were "as likely as not" to choose Punish? If so, then we will assign to the outcome $1 a utility number less than 50 (where we assign 0 to $0, 100 to $100) If he would still prefer Resist if he thought "there was a 3-1 chance" the blackmailer would pick Punish, ~~we~~ the utility of $1 would be less than 25. ~~With~~ What if he thought it "90% likely" that the Blackmailer would choose Punish? If we decide that he would still choose Resist, we have restricted the utility of $1 to a number below 10: which may be as~~fx~~ far as our intuition will take us. It is a useful distance; and far enough to distinguish this case from the preceding one.

Thus, we can _ask_ questions about "critical risks" prior to any measurement of utilities; and to the extent that we can answer them, we _get_ estimates of the utilities. Conversely, if we can make even rough estimates (or guesses) as to his utility payoffs (i.e., as to his preferences among lottery tickets offering these outcomes) we can deduce, roughly, a "critical risk" for him. ~~inthese~~ In the above case, we might estimate the critical risk simply as _high_; the victim

must be "very sure" that the threatened punishment will be carried out before he will comply with the blackmailer's demand. This crude ~~rough~~ information, derived from the "objective" outcomes with the aid of introspection and practical experience, may be quite enough: a) to contrast this game usefully with others (in which, for example, the critical risk is clearly <u>low</u>); b) to support a prediction as to the victim's choice in this game, even when his actual subjective probabilities are known only roughly (for example, if we estimate that he regards the "actual risk" of punishment in this case as <u>low</u>).

Thus, the formulation can exploit crude data. From now on, for convenience in exposition, we shall assume that the von Neumann-Morgenstern utility payoffs are known precisely by us. This is far from realistic, and we must keep in mind that in an application using empirical data (whether estimated introspectively or through controlled experiments), a comparison of "critical risk" with "actual risk" will lead to reliable conclusions only when these variables ~~are~~ show gross, obvious differences.

Given precise payoffs, it is time-saving to know a simple algorithm for computing the "critical risk." ~~Given~~ Consider a 2x2 matrix in the following form, where the payoffs refer to the column player:

| r | Comply | Resist |
|---|---|---|
| Accept | $U_{11}$ | $U_{12}$ |
| Punish | $U_{21}$ | $U_{22}$ |

1) First, check to see that neither column is dominant. 2) If neither strategy dominates, then subtract the first column from the second

column, and set the results in a new column, <u>in inverted order</u> (i.e. the difference for the first row on the bottom, the difference for the second row on top). Minus signs can be disregarded; only absolute differences are relevant. Thus:

$$\begin{array}{cc|c} U_{11} & U_{12} & |U_{22} - U_{21}| \\ U_{21} & U_{22} & |U_{12} - U_{11}| \end{array}$$

These numbers correspond to what J.D. Williams has termed "oddments";[15]

15. <u>The Compleat Strategyst</u>, New York, 1954, p. 40.

or fragments of odds. ~~They~~ Compared to each other, they represent the subjective odds on the opponent's (in this case, blackmailer's) strategies that would make the victim indifferent between his own two strategies (thus, the "critical odds"). (E.g., 2:2, 3:1, 7:5, etc.
3) To convert these odds into probabilities, divide each "oddment" by the sum of the "oddments."

$$\left[\begin{array}{l} |U_{22} - U_{21}| \,/\, |U_{22} - U_{21}| + |U_{12} - U_{11}| \\ \\ |U_{12} - U_{11}| \,/\, |U_{12} - U_{11}| + |U_{22} - U_{21}| \end{array}\right.$$

These subjective probabilities are "critical" in the sense that, as "actual" subjective probabilities they would leave the player indifferent between his alternative strategies. The lower figure, corresponding to the opponent's Punishment strategy, is what we have termed the victim's "critical risk". If he assigns a higher probability to Punishment, he will choose to Comply; if lower, to Resist.

With this method, one can write down the critical risk almost as fast as the payoff numbers themselves. The steps are shown in this numerical example: (notice that the signs of the differences are disregarded):[16]

| | Comply | Resist | |
|---|---|---|---|
| Accept | 25 | 100 | 75/100 |
| Punish | 75 | 0 | 25/100 |

We have not assumed in this example that the first column consists of a constant, an assumption we have been using so far for convenience.

16. The simple arithmetic underlying this procedure is as follows. We are looking for a probability p for Punishment such that the victim is indifferent between Comply and Resist; the following equation determines p:

$$(1-p)\,U_{11} + p\cdot U_{21} = (1-p)\,U_{12} + p\cdot U_{22}$$

$$\text{so:}\quad p = \frac{U_{12} - U_{11}}{(U_{12} - U_{11}) + (U_{21} - U_{22})}$$

Let us now return to our formulation of the blackmailer's problem: to ensure that the victim estimates his actual risk as higher than his critical risk. (In the above numerical example, this becomes: to ensure that the victim assigns more than 1/4 probability to Punish). A great many diverse patterns of behavior can be understood very simply from this point of view: as threat-behavior, designed to increase a victim's estimate of his actual risk relative to his estimate of his critical risk.

By his very choice of threat, of course, the blackmailer can influence his opponent's payoffs, his critical risk (i.e., when several alternative actions are available to the blackmailer as threats). Other things being equal (which will rarely be the case), the lower the victim's critical risk, the easier the task of the blackmailer. Again, for given strategies and given objective outcomes, the blackmailer may aim to influence the victim's preference hence his payoffs and critical risk. Or he may seek to change the victim's expectation ("perception") of the objective consequences of a given pair of opposing strategies. All of these have the effect of changing the victim's utility payoffs. For example, given that the victim initially sees his payoffs as:

| Comply | Resist |
|---|---|
| 50 | 100 |
| 50 | 0 |

the blackmailer may attempt to change his perception of the payoffs to:

| | |
|---|---|
| 50 | 100 |
| 50 | -150 |

or

| | |
|---|---|
| 50 | 625 |
| -50 | 0 |

or

| | |
|---|---|
| 80 | 100 |
| 80 | 0 |

Any one of these modifications changes the victim's critical risk from 1/2 to 1/5; instead of making his threatened action appear "~~as likely as not" to~~ "more likely than not" to the victim, he need convince the victim only that there is a "20% chance"--i.e., "a small chance" that the threat will be carried out. If in the

course of effecting this transformation the blackmailer has not simultaneously made his punishment strategy appear much less likely than before, he must now be closer to success. In fact, if the victim's expectations (his estimate of the "actual risk") have remained constant, and if he already assigned more than 20% likelihood to Punish, the blackmailer's problem is solved.

In general, we have assumed ~~that~~ and will continue to assume that the opponent's payoffs are regarded as fixed. However, a good deal of bargaining behavior is undoubtedly aimed at changing these payoffs, and non-game-theoretical discussions of bargaining have often emphasized such behavior.[17] Though we have not attempted here

17. See Stevens, Penn, etc.... Dunlop...Lindblom.
In this category is such behavior as a union's attempt to convince an employer of the losses to be suffered in a strike, or to educate him as to benefits to him of paying a higher wage rate (in higher productivity, better worker morale, public opinion, freedom from government interference, increased purchasing power of workers, etc.).

to formalize the opportunities for such changes or the process for achieving, the comments above suggest that it is easy enough in our simple model to represent the final results of this process and their significance for the players. From now on, we will assume that the strategies in the game and the victim's payoffs are given.

The critical risk being given, the blackmailer must estimate the victim's "actual risk," i.e., his actual expectations, and if necessa modify ~~influence~~ these so that the victim will regard the risks associated with resistance as "too great." How does he set out to influence the victim's expectations; what are the obstacles, what are the limitatio on his ability to do this? These are the central problems in blackma

Let us return, for convenience, to the notion that "I" am the blackmailer, "you," the victim. How, then, do you form your expectati

E

of my behavior in the first place; how would you arrive at them if I were doing nothing to influence you?

You could go by past experience, if you knew any that seemed relevant. How have I behaved in games with others? With you? Do I "tend" to carry out my threats, my predictions, or do I bluff? How do people "like me" typically behave, according to the papers or your experience? Such information is "external" to the game as defined, and orthodox game theory has entirely ignored it, since there is no place for it in the standard game analysis, no element in the game situation as defined which ~~it~~ it could be held to influence. But since the present analysis includes expectations as an explicit variable, such information as this will naturally affect the victim's (or in general, a player's) subjective probabilities; and assumptions as to the information available to the opponent will lead to inference about his expectations, hence will influence one's own expectations.

Thus we can easily allow for the ~~fa~~ influence in the forming and estimating of expectations of factors that lie beyond the elements that define the game, the strategies and payoffs. Presently we shall consider the important influence on expectations of the payoffs, but ~~there is no~~ it is a starting point of this analysis that payoffs are not the sole, determining, influence on expectations. The other influences are so various (including not only experience, dogma, prejudice, authority, but uncertainty as to the opponent's payoffs and <u>his</u> expectations) that it seems most convenient to represent them implicitly through their ~~infln~~ effect on the "independent" variable, expectations.

Thus, if I know "the record" you are likely to be looking at, I can guess at some aspects of your expectations. However, few of the influences mentioned are relevant to my problem of <u>changing</u> your expectations (though they may determine how hard it will be to change them). I may be able to fill in the record a bit for you, underline certain parts, lie about others, "correct" certain "errors"; but on the whole there is little I can do to manipulate your expectations by changing the record.

But with or without experience to go on, the victim has another basis for expectations: his knowledge of the <u>blackmailer's payoffs</u>. If you regard me as "reasonable," then your question, "How is he likely to respond to my actions?" becomes: "What will he <u>want</u> to do? What will it pay him to do? Which will he see as the <u>best</u> choice for him to make, <u>given</u> my choice?"[18] These questions call for knowledge

18. Recall that in our model of the blackmail situation, it is assumed that the victim will choose first, knowing that the blackmailer will choose in full knowledge of the victim's choice.

of the ~~p~~ "objective" outcomes to me that (I expect) will result from ~~a~~ given pairs of opposing strategies, and of my preferences for these outcomes: i.e., my payoffs.[19]

19. It does not immediately follow that you must know more than my ordinal preferences over these outcomes; but in fact, we shall see that my von Neumann-Morgenstern utilities are also relevant.

You may have no idea what my payoffs are, in which case these questions go unanswered. At best, your estimates will be rough and uncertain, perhaps even more than my estimates of your payoffs (since it is common for a blackmailer--in the familiar sense--to know a good deal more about his victim than his victim knows about him). Still,

you usually will make some assumption, perhaps very well informed, about my payoffs; and when you do, it will influence your expectations.

And this can be the blackmailer's primary obstacle, the crucial limitation on his ability to influence. The effect of his own payoffs on the victim's expectations will often be the factor that he must change, counteract, surmount, if he is to succeed. For typically, if the victim should fail to "obey," it would be <u>costly</u> for the blackmailer to carry out his threatened punishment. It would not, in general, give him his best outcome under the circumstances: i.e., given the victim's actual, rebellious choice. Which is to say, it would be "irrational" to carry out the threat. It would mean, for the blackmailer, deliberately passing up an outcome that he preferred; he could, if he chose, do better. Why wouldn't he? That is the thought that is bound to occur to the victim; and the blackmailer knows it.

In our diagram we need only add the blackmailer's payoffs, as numbers (we shall assume these are expressed as von Neumann-Morgenstern utilities) to the left of the victim's payoffs; from now on, the first number in the box at the intersection of a row and column in the matrix signifies the payoff to the "row" player, the second the payoff to the "column" player.

| | Comply | Resist |
|---|---|---|
| Accept | 100, 30 | 50, 100 |
| Punish | 100, 30 | 0, 0 |

In the above example, we assume that the blackmailer has threatened to choose his second strategy, Punish, if the victim chooses his second strategy, Resist. If the victim does choose Resist and the

threat is carried out, the victim gets a payoff of 0, instead of the 30 that he could have received with certainty for choosing Comply. But to carry out the threat is costly also for the ~~the~~ blackmailer; his payoff, from choosing Punish, is 0 instead of the the 50 that he could have won, with certainty (given the victim's prior choice) by choosing Accept. Can he convince the victim that, if the occasion arose, he would pick the 0 instead of the 50?

The first point to realize--~~th~~following directly from our eariler discussion--is that he need not make the victim certain of this; the victim must believe merely that the threatened action is "sufficiently likely" to make the strategy Resist "too risky." Knowing his payoffs, we can give a precise meaning ~~to (if not a precise value)~~ to "sufficiently likely"; it means a likelihood higher than the critical risk. The question we now face is: how does the blackmailer make his threat appear even slightly likely, against the evidence of his own payoffs? It is not the exception but the rule for the threat to have a certain built-in implausibility, being costly--or irrational--for the threatener to carry out. His efforts to overcome this barrier to belief account for the most characteristic, and paradoxical, features of threat-behavior.

To make it plausible that he will carry out a costly threatened action, a threatener has four main approaches. The first two, in particular, have been analyzed and illustrated exhaustively in Thomas Schelling's brilliant article, "An Essay on Bargaining,"[20]

20. American Economic Review, June, 1956...

to which I refer the reader for extended discussion. First, the blackmailer can voluntarily but irreversibly give up his freedom of choice; he can make it impossible for himself not to carry out

HK—DM ("We may have done this. [illegible]

TCS: threats that leave us make it sufficiently like



his threat. If he can in some way bind his own hands, destroy his alternatives, he may be able to "make it true" that he would carry out his threat, for the reason that he would have no choice.

This might be represented in our model by his deleting certain rows, or entries, from the payoff matrix, symbolising: a) that these have been eliminated as alternatives for him; b) that his opponent recognizes this (where this is not strictly true, this representation would be inappropriate). Thus, in our last example, the blackmailer might be able to "strike out" the first row, perhaps thus "compelling" the victim to believe that the consequence of his Resist strategy will be 0.

| | |
|---|---|
| ~~100, 30~~ | ~~50, 100~~ |
| 100, 30 | 0, 0 |

Thus we find the blackmailer striving to achieve his goal by eliminating ~~alternatives~~ opportunities, contracting his set of alternatives, althou[gh] discussions of bargaining often rule out such behavior axiomatically.[21]

21. See Nash...

Our explanation is that he hopes for a favorable effect on his opponent's expectations (a goal bound to be ~~ignored~~ eliminated by treatments that ignore expectations).

~~However,~~ Actually, it may be rather rarely that a blackmailer can ~~might~~ irrevocably "tie his hands"/or destroy alternatives quite literally. But even if he cannot make given actions (i.e., failure to punish) impossible, he may be able to make them costly. Although Schelling

does not emphasize the distinction, most of the examples in his essay of "commitment" (which might also include the above tactics) fall into this ~~xx~~ category. The player binds himself to incur certain costs or penalties or forego certain advantages if he should fail to carry out a pledge; thus he reduces his own payoff incentives to break the pledge (e.g., fail to carry out a threat), hoping thereby to make his pledged action seem more "likely" to his opponent. We can represent this behavior very neatly in our formal model by allowing the player specified opportunities to lower his own payoffs; ~~allowing~~ this "move" serves to formalize most of the behavior examined by Schelling, behavior which is ignored or excluded in game-theoretical discussion and which appears puzzling or perverse in actual experience.[22]

22. I suggested this formalization of Schelling's treatment of committment in a seminar at the RAND Corporation, in the summer of 1958. at which TCS was present—and where no contradiction it was not novel to him. Schelling's most recent analysis, shows that he had arrived independently at precisely the same approach.

CLAIMS.

Perhaps the blackmailer can make a contractual agreement with a third party to choose Punish if the victim chooses Resist, making himself liable to forfeit, penalty, or suit if he fails to carry out this action. Or he can stake his honor, his prestige, his reputation for honesty--if he has any--on carrying out this prediction. ~~(Thus, xxxxx~~ These new obligations can be symbolized readily in our model as subtractions from ~~the~~ his previously-estimated payoff to choosing Accept when the victim chooses Resist (i.e., failing to choose Punish). If he can actually lower that payoff below that for Punish, he will have removed his evident incentive to back down from his threat, which presumably will become much more convincing to his

opponent.

Thus, with these payoffs:

| | |
|---|---|
| 100, 20 | 10, 100 |
| 100, 20 | 0, 0 |

he may seek to change his own payoffs in this fashion:

| | |
|---|---|
| 100, 20 | -1, 100 |
| 100, 20 | 0, 0 |

He has done nothing but worsen one of his own payoffs; yet he may have improved considerably his chances for a favorable outcome from the play. He has made it "plausible" that he will carry out his threat: by making it "rational" to do so.[22]

22. It would seem natural to assume from this that the victim is subsequently certain to Comply; but there are good reasons for avoiding this assumption, ~~which~~ discussed below.

It might appear that such a "move" should properly be included in the formal ~~game~~ strategies open to the player, defining the game. That would simply amount to a different abstraction of the empirical situation; I believe it is less useful to obscure the peculiar nature of this particular tactic, the reduction of the payoffs to certain actions for the purpose of influencing the opponent's expectations.

It should be noted that a given penalty or forfeit, stated in money or other objective terms, will mean different reductions in "utilities" depending on the outcome with which it is associated. (I.e., unless utility is linear on money, a reduction of outcome from $100 to $0 will not have the same effect on utilities as a reduction

from $1000 to $900.) Also, players will differ greatly in their ability to reduce their own payoffs. Honor, prestige, reputation are useful to the blackmailer (as to the promiser) because they can be pawned. They can be wagered, risked, pledged as security; they represent something worth preserving, which can make more credible choices designed to preserve them. But their worth (hence, their usefulness as security) will vary from player to player. The man without honor, with a poor record for honesty, with little prestige, has little to lose if he breaks his prediction; he would not be expected to resist strong temptation. His solemn promise to carry out his threat might, in the eyes of his opponent in the example above, lower his payoff only moderately, ~~withxinconclusi~~

SMM on resigning from Friend Andrew

| | |
|---|---|
| 100,20 | 5, 100 |
| 100,20 | 0, 0 |

The effect of this change on the victim's expectations might be inconclusive. Similarly, players may be differently situated in their ability to undertake contractual commitments, or to convince their opponents that these undertakings would be enforced. Finally, a given player will have differing ability to lower his payoffs in different games, ~~with~~ in different institutional environments, and perhaps ~~with~~ in connection with different strategies.

To describe, then, the opportunities for commitment (taking this to cover tactics that correspond to deleting strategies or to lowering payoffs) open to a player, we would have to list all the transformations of his utility payoff matrix that he could effect.

It might be possible to reflect gross differences in the ability to commit oneself ~~by~~ specifying, in connection with a given utility scale, a maximum amount which a given player can subtract from any payoff. This amount might be "enough" for the player's purposes in one game, "not enough" in another (depending on the amount of "temptation" the commitment must overcome). Or in a given game, the tactic of commitment might appear quite promising for ~~the~~ a player who can make "large commitments", but not for another who cannot reduce payoffs so appreciably.

A third general approach to the problem of enlarging the opponent's "actual risk" above his critical risk is to create and explicit uncertainty in the victim's mind as to the blackmailer's "true" payoffs (i.e., the blackmailer's perception of his own payoffs). Thus the victim becomes unsure in his predictions of ~~the~~ what actions ~~that~~ the blackmailer actually regards as rational. ~~(In game theory, of course, this tactic is ruled out by the assumption~~

Finally, the blackmailer can strive to appear irrational: for which purpose it may be helpful for him to be irrational. Irrational, perhaps, in being erratic, inconsistent, unpredictable, on the basis of his known, independently determined preferences for the outcomes expected. (Choices not in accord with these ~~pref~~ assumed preferences might be regarded as calling these preferences into question; what I am suggesting here is that the blackmailer may try to give the impression that though he has known, well-defined preferences for the outcomes, reflected in his free and considered choices, his actions may occasionally be governed by impulse, unconscious hidden constraints, (ID) authorities, influences, or by insubordinate agents).

Both these last two classes (relatively ignored by Schelling) ? might seem too bizarre to deserve formal consideration. Moreover, as we shall see later, they call into question the assumption that the victim assigns subjective probabilities to the blackmailer's choices. Nevertheless these tactics occupy in fact far too large a role in actual blackmail and bargaining behavior to be ignored. It might be noted that both tactics (creating uncertainty as to one's ~~truth~~ payoffs, and creating uncertainty that one's choice will be consistent with the payoffs) are particularly promising when the opponent's critical risk is very low. They suggest to the opponent that there is at least a small chance that the blackmailer will ~~be~~ do "almost anything," even carry out what appears to be a very costly threat.

| | Comply | Resist |
|---|---|---|
| Accept | 100, 99 | 1, 100 |
| Punish | 100, 99 | 0, 0 |

The victim's critical risk in the above game is 1/100; to win, the blackmailer must ensure that the victim assigns at least 1% likelihood to the strategy Punish if he picks Resist. If the blackmailer could "commit" himself in the ways we considered earlier, by deleting his ~~second~~ first strategy, or by reducing his payoff to (Accept, Resist) from 1 to -1, he would probably accomplish this. These are, I would say, the tactics that Schelling stresses in his "Essay on Bargaining." Yet neither might be available to the blackmailer; he might have no means to bind himself irrevocably, or to reduce his payoffs unmistakably.

Even though he can neither "make it true" that, given Resist, his payoff to Punish is better than his payoff to Accept (e.g., 0 compared to -1) nor make the victim attach high likelihood to this state of affairs, he may still be able to persuade the victim that such payoffs are ~~x~~ "possible." If ~~man~~ he can convince the victim that a haze surrounds all the payoffs, that none of these estimates is to be trusted implicitly, the victim might conclude that there is "some chance" that the upper right-hand payoff to the blackmailer is "really" -1. If he, in effect, assigns as much as 1% probability to this possibility, the blackmailer wins. (Cheney's 1% rule!)

The usefulness of this tactic clearly depends on the fact that the victim must be made only a little uncertain of the payoffs; i.e., the victim will Comply if he assigns even a small likelihood to the possibility that the blackmailer "really" prefers Punish. Even so, this uncertainty will be bounded. With a given utility scale, if the victim's "best guess" of the blackmailer's utility for a given outcome is 0, it should be easier to convince him that it ~~is~~ may "really" be 1 than that it may really be 100. (i.e., that the ~~haze~~ range of the haze, or uncertainty, around each payoff is 2-3 "utils," not 100 "utils." But that is not to say that it will be easy; we can't say a priori that the utility difference ~~is small~~ between 0 and 1 is small (only that it is smaller than between 0 and 100), or that an error of this size is "likely" or "plausible." The unit and origin of our utility scale are arbitrary, and it might be very difficult or impossible to convince the victim that there was even as much as 1% possibility of an error as large as 0-1. All we can say is that this tactic of creating doubt as to the payoffs (we have said nothing about the method

for achieving this) ~~it~~ seems more promising in the above case than in the following one, where the same utility scale is assumed:

| | |
|---|---|
| 100, 1 | 99, 100 |
| 10, 1 | 0, 0 |

Here the victim's critical risk is .99. Even if the blackmailer's payoffs were the same as before, he would ~~probably~~ find it impossible to induce the victim to Comply just by creating uncertainty as to his own payoffs. (Making the victim more than 99% sure that the blackmailer would prefer to Punish is, in effect, changing thepayoffs for him; it is not "creating uncertainty"). Moreover, if the victim's payoffs had remained the same as before, we have assumed new payoffs for the blackmailer such that an error of the relevant size must seem much less likely to the victim. ~~In other words:~~ The victim might not attach even as much as 1% probability to the possibility that the blackmailer's payoff of 99 is "really" -1 (if, before, he could just barely be induced to attach 1% likelihood to the possibility that 0 was "really" -1).

The last two examples also illustrate the conditions which favor or disfavor our last tactic, "irrationality." Suppose that in the previous example, there is no hope of clouding the victim's perception of the payoffs. He knows, with certainty, that faced with the victim's choice of Resist, the blackmailer would prefer the outcome to Accept to the outcome for Punish. He must be persuaded that, *nevertheless*, the Blackmailer *might* choose Punish. ~~(A~~

Again, it is the victim's low critical risk that makes the outlook at all hopeful; he does not have to find Punish *very* likely, before he chooses Comply. If the blackmailer ~~EXXXE~~ displays (in the victim's mind) a certain randomness in behavior--if his behavior is not *entirely* predictable from his known payoffs--the victim may conclu that Resist is too risky. We assume that this erratic behavior is not based upon uncertainties or shifts in his payoffs; as mentioned earlier, it could reflect the intermittent influence of authorities or agents.[24]

24. The term "irrational" refers to inconsistencies in his observed behavior; if we *knew* some explanation for them--such as the possibilities mentioned--we would use some other term. Our usage here is based on the assumption that he has known expectations of outcomes and known, stable and consistent preferences, but that for some reason his choice does not accord with these.

OR WEIRD, "MAD", PRIORITIES! HITLER!

The blackmailer might even "make it true" that he *might* pick Punish by adopting, in effect, a mixed strategy: openly, *before* the victim chooses. Abstractly, he would be relinguishing his free choice after the victim's move to a random mechanism; ~~withxfixedxprobabilities~~; he would be adopting probability constraints on his own response. His motive would be to influence the victim's expectations; as with the other tactics, we will not assume that the results of his maneuver on his opponent's expectations are entirely predictable. In particula we need not assume that the victim subsequently attaches the same subjective probabilities to his strategies that he actually attaches in his mixed strategy. By making ~~hisxchoicexof~~ Punish "actually" 10%, or 20%, or 50% likely, the blackmailer would hope to induce the

the blackmailer

victim to regard it as ~~xxtrlxx~~ "more than 1%" likely. (The probability constraints on his behavior might not specify an exact probability distribution; they might, in fact, ensure merely, a probability for Punish "greater than 1%, 10%, etc.")

Of course, given that the payoffs are von Neumann-Morgenstern utilities, to choose one's strategies with given probabilities is ~~inn~~ in effect to choose a mixed strategy, with its own expected payoffs for given choices by the opponent. For example, the payoffs ~~xxxxxx xxxxxxxxxx~~ of a mixed strategy combining ~~xx~~ Accept with 9/10 probability and Punish with 1/10 probability ~~xxx~~ can be shown:

| | | |
|---|---|---|
| | 100, 99 | 1, 100 |
| | 10, 99 | 0, 0 |
| (.9, .1) | 91, 99 | .9, 90 |

(The victim's payoffs here reflect the assumption that the victim assigns the "correct" subjective probabilities to the blackmailer's pure strategies). In effect, the blackmailer chooses first: a mixed strategy. He can now expect that the victim will choose Comply, to get an expectation of 99 instead of 90. He himself can no longer achieve an expectation of 100 (since, we assume, he has "committed" himself to a random mechanism which has a 10% chance of choosing Punish), but he can have an expectation of 91, which is far better than the 1 he would expect if he made no attempt to influence the victim's choice.

Even better for the blackmailer would be a "contingent" mixed strategy, which would commit him to some probability of Punish if the opponent chose Resist, but which would allow him to Accept if the

victim choose C only.

This might be regarded as a generalization of the first form of commitment, which amounted to the choice of a mixed strategy with probability 1 for one ~~xx~~ pattern of response and 0 for all others. When we take the victim's critical risk into account, such an extreme commitment may seem unnecessary; it might be just as effective, and perhaps more feasible or less costly to introduce a moderate amount of randomness into one's choice. A history of somewhat erratic choice in past games (where, in fact, the stakes may have been much lower), a tendency to yield to impulse, suggestion or rage, occasional unreliability of one's subordinates, a reputation or appearance of "madness": all these could be useful assets in a particul game (even though--<u>if true</u>--they may limit the maximum that may be achieved).

These last two classes of tactics are particularly important in cases such that: a) the victim has a low critical risk, but (b) the threatened action would "normally" seem to entail such costs for the blackmailer that a "normal," rational blackmailer would find it difficult to make his threat even slightly credible (see the discussion <u>below</u> of nuclear deterrence and bank robbery). In other words, the cost to a "normal" or typical threatener of carrying out the threat (say, nuclear devastation) or in general, threats of suicidal action) appears so great that such a person could not actually "make it rational" by processes of commitment of the sort that Schelling describes. It may then help to appear "mad," in one of two distinct senses: a) having (at least, possibly) "mad" or "abnormal" preferences, e.g., some taste ~~nf~~ for suicide, so that the threatened action appears less costly than it would be for the normal person; b) having, perhaps, normal payoffs for the outcomes concerned, but not being

governed wholly in one's action by rational considerations.[25]

25. A study of Hitler's ultimatums, e.g., on the occasions of the Anschluss, Munich, and the occupation of Prague (also, implicitly, the occupation of the Rhineland), suggests that his success depended heavily on his reputation for both these forms of "madness."

It may be difficult to make an opponent _certain_ that one's preferences are wildly abnormal or that one's behavior is wholly erratic; hence the significance of the first condition, that the victim have a low critical risk (so that he doesn't need to be sure).